REVIEW 10.1111/1469-0691.12056

Next-generation sequencing technology in clinical virology

M. R. Capobianchi, E. Giombini and G. Rozera

National Institute for Infectious Diseases 'L. Spallanzani', Rome, Italy

Abstract

Recent advances in nucleic acid sequencing technologies, referred to as 'next-generation' sequencing (NGS), have produced a true revolution and opened new perspectives for research and diagnostic applications, owing to the high speed and throughput of data generation. So far, NGS has been applied to metagenomics-based strategies for the discovery of novel viruses and the characterization of viral communities. Additional applications include whole viral genome sequencing, detection of viral genome variability, and the study of viral dynamics. These applications are particularly suitable for viruses such as human immunodeficiency virus, hepatitis B virus, and hepatitis C virus, whose error-prone replication machinery, combined with the high replication rate, results, in each infected individual, in the formation of many genetically related viral variants referred to as quasi-species. The viral quasi-species, in turn, represents the substrate for the selective pressure exerted by the immune system or by antiviral drugs. With traditional approaches, it is difficult to detect and quantify minority genomes present in viral quasi-species that, in fact, may have biological and clinical relevance. NGS provides, for each patient, a dataset of clonal sequences that is some order of magnitude higher than those obtained with conventional approaches. Hence, NGS is an extremely powerful tool with which to investigate previously inaccessible aspects of viral dynamics, such as the contribution of different viral reservoirs to replicating virus in the course of the natural history of the infection, co-receptor usage in minority viral populations harboured by different cell lineages, the dynamics of development of drug resistance, and the re-emergence of hidden genomes after treatment interruptions. The diagnostic application of NGS is just around the corner.

Keywords: Co-receptor usage, metagenomics, next-generation sequencing, quasi-species, resistance-associated mutations, ultradeep pyrosequencing, viral diversity

Article published online: 3 October 2012

Clin Microbiol Infect 2013; **19:** 15–22

Corresponding author: M. R. Capobianchi, National Institute of Infectious Diseases INMI, Lazzaro Spallanzani, Via Portuense, 292-00149, Rome, Italy

E-mail: maria.capobianchi@inmi.it

Next-generation Sequencing: Principles and Methods

Recent advances in the nucleic acid sequencing technologies, referred to as 'next-generation' sequencing (NGS), have produced a true revolution and opened new perspectives for research and diagnostic applications. One of the hallmark features of the NGS technologies is their massive throughput at a modest cost, with hundreds of gigabases of sequencing now being possible in a single run for some thousand dollars.

The first commercially available NGS system, developed by 454 Life Sciences, appeared in 2005. Since then, in a relatively short time, several NGS technologies have been developed (Table 1). According to the different characteristics, the spectrum of application of the platforms may show significant differences. In particular, the systems can be referred to as: high-capacity sequencers, such as the Genome Analyzer, HiSeq sequencers (Illumina, San Diego, CA, USA), the Heliscope platform (Helicos BioSciences Corporation, Cambridge, MA, USA) and 5500 series sequencers, which use SOLiD technology (Applied Biosystems, Carlsbad, CA, USA); and long-read sequencers, such as the Genome Sequencer (GS) FLX or Junior (454 Life Sciences, Roche Diagnostics, Branford, CT, USA), Ion Torrent (Applied Biosystems), and PacBio RS (Pacific Biosciences, Gen-Probe, Menlo Park, CA, USA).

Current NGS methods use a three-step sequencing process: library preparation, DNA capture and enrichment, and sequencing/detection [1]. In library preparation, DNA (or cDNA) fragments of appropriate lengths are prepared, by

TABLE 1. Characteristics and principal applications of next-generation sequencing platforms

Manufacturer	Platform	Maximum throughput per run (Gb)	Maximum sequence length (bp)	Template preparation/sequencing	Principal applications	Error source
Illumina	Hi Seq 2000	600	100	Solid capture/bridge amplification *in situ*/reversible chain terminators	Genome resequencing, quantitative transcriptomics, genotyping, metagenomics	Signal interference among neighbouring clusters, homopolymers, phasing, nucleotide labelling, amplification, low coverage of AT-rich regions
Applied Biosystems	SOLiD	15	75	emPCR/ligation and two-base coding	Genome resequencing, quantitative transcriptomics, genotyping	Signal interference among neighbours, phasing, nucleotide labelling, signal degradation, mixed beads, low coverage of AT-rich regions
	Ion Torrent PGM	1	200	emPCR/real-time sequencing with detection of H^+	*De novo* genome sequencing and resequencing, target resequencing, genotyping, RNA sequencing on low-complexity transcriptome, metagenomics	Homopolymers, amplification
Pacific Biosciences/Gen-Probe	PacBio RS	0.045	1200	Single-molecule/linear amplification, real-time sequencing, fluorescent nucleotides	*De novo* genome sequencing and target resequencing, non-amplifiable samples, PCR-free	Low intensities
Helicos	Heliscope	35	35	Single-molecule/reversible chain terminators	Direct RNA sequencing, non-amplifiable samples, PCR-free, and unbiased quantitative analyses	Polymerase, molecule loss, low intensities
454 Life Sciences/Roche Diagnostics	GS FLX+	0.7	1000	emPCR/pyrosequencing	*De novo* genome sequencing and resequencing, target resequencing, genotyping, metagenomics	Homopolymers, signal cross-talk interference among neighbours, amplification, mixed beads
	GS Junior	0.035	400		Target resequencing (amplicons), genotyping	

emPCR, emulsion PCR.

either breaking long molecules, or by synthetically preparing short molecules (i.e. by PCR or cloning). In the DNA capture and enrichment phase, these short molecules are labelled with primers that are used to capture and physically separate each single short fragment, fixing it onto a solid substrate. Each single molecule acts as a template for clonal amplification (single-molecule template principle). The sequencing phase is based on DNA polymerization combined with detection. These steps occur concomitantly on myriads of clonally amplified fragments. In some of the platforms, the PCR step is not required, and sequencing is performed on single (Helicos) or on linearly amplified (PacBio RS) molecules. The principles for the sequencing/detection phase are different in the various platforms: pyrophosphate release (pyrosequencing) coupled with optical detection of fluorescence, hydrogen ion release coupled with detection of pH variation by a semiconductor, ligation combined with fluorescence detection, and linear amplification coupled with fluorescence detection. The detected signals are transformed into sequences by an integrated browser and elaborated with specific software, to match pre-established quality scores. Bioinformatic tools for the evaluation and elaboration of the sequence data are in continuous development, according to the various applications.

There are many factors involved in the choice of technology, including cost performance, run time, accessibility, type of application, and convenience. More detailed descriptions of the technical features of NGS methods are given in [2–6].

NGS Applications to Virology

NGS applications to virology have been recently reviewed [7–10]. Some of the topics relevant for clinical application will be briefly described.

Virus discovery (metagenomics)

The term metagenomics designates the analysis of all of the nucleic acid present in a given sample, allowing the exploration of entire communities of microorganisms, and avoiding the need to isolate and culture individual microbial species, and does not need previous knowledge of the sequences.

This new science is one of the fastest advancing fields in biology, and is extending our comprehension of the diversity, ecology, evolution and functioning of the microbial world, as well as contributing to the emergence of new applications in many different areas. Several large projects have been funded in the USA or EU for the definition of the human microbiome based on metagenomics (e.g. the Human Microbiome Project,